Statistical Analysis Plan

INTERVENTION	Achieve Together
DEVELOPER	Future Leaders, Teach First, Teaching Leaders
EVALUATOR	Institute for Fiscal Studies
TRIAL REGISTRATION NUMBER	Not registered as not a randomised controlled trial.
TRIAL STATISTICIAN	Ellen Greaves
TRIAL CHIEF INVESTIGATOR	Ellen Greaves
SAP AUTHOR	Ellen Greaves
SAP VERSION	4
SAP VERSION DATE	09/08/2016
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Table of contents

Introduction

Achieve Together is an initiative designed and delivered by three education charities - Teach First, Teaching Leaders and Future Leaders, with the ultimate aim of raising academic standards in schools in low-income communities. The core part of Achieve Together is the presence of these three programmes in the same school. Future Leaders develops existing senior members of staff through training programmes that support aspiring head teachers and the practice of senior leaders, or places senior leaders in the school. Teaching Leaders provides specialist training to promising middle leaders, which includes coaching and support, formal training outside school hours, and an 'impact project'. Teach First places carefully selected trainee teachers in schools, usually to teach high-demand subjects. These trainees work towards qualified teacher status (QTS) and a PGCE qualification while teaching and being paid a salary. Achieve Together therefore incorporates the provision of new high-potential teachers and senior leaders as demand requires, and training for existing staff. Achieve Together also sought to bring these programmes together through a collective impact project and various activities designed to increase alignment between the individual leadership development training and whole school improvement priorities. The evaluation will provide evidence on whether this intensive investment in the human capital of teachers in a school leads to an improvement in pupils' academic performance.

The impact of the intervention on pupil attainment will be measured by their GCSE capped average points score at the end of the two year intervention. It will be possible to measure the longer-term impact of the intervention in later years.

A well-matched comparison group will be used to provide a counterfactual for attainment in the intervention schools, as an insufficient number of schools were recruited for the original randomised controlled trial proposed.

Study design

The population of intervention schools is the 14 schools that signed up to the Achieve Together trial (beginning in September 2014) with full knowledge of the original randomised controlled trial design. These schools were selected from schools that met the eligibility criteria used by Teach First, Teaching Leaders and Future Leaders at the time: at least 50% of pupils must come from the lowest 30% of the IDACI (Income Deprivation Affecting Children Index); or at least 50% of pupils are eligible for free school meals, or performance at KS4 is below the national average for schools where the proportion of pupils eligible for free school meals is between 25% and 50%. An additional eligibility criterion was that the school could not currently be working with all three partner charities (but could have been working with one or two). Schools that had existing engagement with one or more charity could potentially have a different treatment effect, but unfortunately this is not possible to explore given the small sample size and availability of information. Schools in London were excluded from recruitment given the existing high engagement with either Teach First, Teaching Leaders or Future Leaders.

The design of this evaluation has changed in response to the number of schools recruited to the Achieve Together Trial. The initial design was a randomised controlled trial, where at

least 48 schools would be recruited and 24 schools would be randomly assigned to the treatment group. The number of schools eventually recruited to the pilot prohibited a randomised controlled trial design, however.

The trial design is instead a well-matched comparison group for the intervention group that received the Achieve Together programme between September 2014 and July 2016. The well-matched comparison group is derived from the group of schools that were eligible for Achieve Together (on the basis of pupil disadvantage) and expressed some interest in Achieve Together before deciding not to participate giving a reason of concern about participating in the trial (as was originally proposed), or financial considerations. These schools are 19% of schools that were eligible for Achieve Together and didn't sign up (13% from financial considerations and 6% from concern about participation in the trial). This group of schools are most likely to have similar characteristics to the intervention group as the Achieve Together programme was of interest to both groups of schools. It is worth noting that deciding not to participate in Achieve Together for either financial reasons or concern about the trial may be correlated with the outcome variables of interest. For example, schools most concerned about participation with the trial might be those that require intervention the most, and therefore have lower pupil outcomes, on average. Similarly, those with financial problems may be expected to have lower pupil outcomes, on average. Alternatively, these schools may have lower pupil outcomes in prior years but be expected to experience higher growth in future years. On balance, however, we think that the benefit of having expressed some interest in Achieve Together (and therefore likely having similar staffing concerns and motivations) outweighs the potential concern that pupil outcomes are systematically different in schools where these reasons for not participating were given. In addition, propensity score matching will be used to select the group of schools with the most similar observable characteristics to the intervention group.

Pupil attainment is measured using the National Pupil Database (NPD) with measurements before the intervention and after the first and second year of the intervention. It will be possible to measure the longer-term impact of the intervention in later years as the well-matched comparison group will not receive Achieve Together at any point.

Protocol changes

The results of the evaluation for the 14 schools that signed up to the Achieve Together trial will not be compared to the results for the 16 schools that actually received Achieve Together between September 2014 and July 2016, as specified in the evaluation protocol. This is because a different matching specification would be required, and differences across the two groups could therefore be due to the matching specification or differences in attainment.

Randomisation

Randomisation did not take place as an insufficient number of schools were recruited to the trial.

Sample size

Initial sample size calculations were completed for the original design of the intervention (randomised controlled trial with 48 schools). These sample size calculations suggested that with a realistic level of intracluster correlation of pupil outcomes within schools (0.2) the minimum detectable effect size would be around 0.28. This assumed 170 pupils per year group in 24 intervention and 24 comparison schools, with desired statistical power of 80% and significance level of 5%, and that the unexplained variance in the outcome variable is 0.6. The equivalent effect size for FSM pupils was 0.29, assuming around 45% of pupils were eligible for FSM in the past six years in intervention and matched comparison schools (based on previous years of NPD data).

This minimum detectable effect size is large relative to existing estimates. The sample size calculations gave a similar minimum detectable effect size once assuming a smaller intervention group and larger comparison group. For the same assumptions of pupils per year per school, statistical power, significance level and unexplained variation in the outcome variable, but with 15 intervention schools and 30 comparison schools the minimum detectable effect size was 0.31. For FSM pupils this was 0.32.

Follow-up

All schools will be included in the final analysis.

Outcome measures

Primary outcome

The primary outcome for the school-level intervention is capped GCSE and equivalents new style point score for KS4 students in Achieve Together schools after two years of the intervention i.e. those that started Year 10 in 2014. This variable will be obtained from the NPD for the full population of pupils in Achieve Together schools in summer 2016 (KS4_PTSCNEWE_PTQ_EE according to the September 2015 edition of NPD Data Tables).[1] This variable will be standardised using the whole population of KS4 pupils to have a mean of zero and standard deviation of one. The sample of pupils used will be common across primary and secondary outcomes. The number of students with at least one outcome missing is highly likely to be very small so these will be excluded from the analysis.

Secondary outcomes

Secondary outcomes are:

1) Capped GCSE point score excluding equivalents (KS4_GPTSCNEWE_PTQ_EE according to the September 2015 edition of NPD Data Tables). This variable will be

[1] https://www.gov.uk/government/publications/national-student-database-user-guide-and-supporting-information

standardised using the whole population of KS4 pupils to have a mean of zero and standard deviation of one.

2) Teach First's measure used for their second "impact goal".[2] This is the capped GCSE point score excluding equivalents plus English and maths bonuses. According to the September 2015 edition of NPD Data Tables this variable would be KS4_GPTSCNEWE_PTQ_EE + Englishbonus_ptq_ee + Mathsbonus_ptq_ee + KS4_ENGLISHBONUS_PTQ_EE + KS4_MATHSBONUS_PTQ_EE. This variable will be standardised using the whole population of KS4 pupils to have a mean of zero and standard deviation of one.

3) Overall absence for KS4 pupils across the academic year (OverallAbsence_6HalfTerms_ab according to the September 2015 edition of NPD Data Tables).

4) Primary and secondary outcomes described above from year one of the intervention (summer 2015).

Analysis

Primary intention-to-treat (ITT) analysis

The analysis model will be at the pupil-level, with robust standard errors clustered at the school-level. A random effects/multi-level model is not used as the main model given the required assumption that the school-level effect is uncorrelated with all pupil and school level covariates in the model. This is not reasonable in this circumstance as pupil characteristics are likely to affect school effectiveness. A sensitivity analysis which includes random effects will be included in the Appendix to the main report, however. The analysis will be intention-to-treat so that all schools, independent of drop-out or engagement with the programme, will be included. As this is a well-matched comparison group design rather than randomised controlled trial pupil- and school-level covariates will be included. These are outlined below. In addition, probability weights derived from the optimal propensity score matching specification will be applied to give the best balance between intervention and well-matched comparison groups. The optimal propensity score matching specification is given below. This was chosen by the specification with the lowest median bias across key school-level covariates after matching. The full set of models tested is given in Appendix 1. This was decided without consideration of the primary and secondary outcome variables, indeed before the primary outcome variable was available.

The model will be run in Stata 14.

Pupil-level covariates:

- Binary variables for season of birth (Winter, Spring, Summer, with Autumn as reference category) from 2016 KS4 NPD data;
- Binary variable for male from 2016 KS4 NPD data;

[2] https://www.teachfirst.org.uk/why-we-exist/what-were-calling

- Binary variables for Ever FSM 6 (yes and missing, with no as reference category) from 2016 KS4 NPD data;
- Continuous variable for IDACI score and binary variable for missing IDACI score from 2016 KS4 NPD data. IDACI score has a large and statistically significant effect over and above FSM in previous years of NPD data;
- Binary variables for attainment at KS2 (lowest quartile, second lowest quartile, second highest quartile and missing, with highest quartile as reference category) from 2016 KS4 NPD data linked to prior attainment at KS2.

School-level covariates:

- Continuous variable for % achieving 5A*-C GCSE and equivalents including English and maths in academic year 2012/2013 from the publicly available performance profiles (ptac5em);
- Continuous variable for total average (capped) point score per pupil (GCSEs only) in academic year 2012/2013 from the publicly available performance profiles (ttapsgcp);
- Continuous variable for value added measure based on the best 8 GCSE and equivalent results in academic year 2012/2013 from the publicly available performance profiles (b8vamea). Value added, total average (capped) points score and %5A*-C have a relatively low correlation for the group of intervention schools which justifies inclusion of all three: between 0.34 and 0.88;
- Continuous variable for % teachers with Qualified Teacher Status in academic year 2012/2013 from publicly available school-level data derived from the School Workforce Census (qts). The theoretical reasoning for this variable is that schools are likely to have signed up to Achieve Together to address staff shortages, which are proxied by this variable;
- Continuous variable for average of latest Ofsted rating for ‘overall effectiveness’, ‘leadership and management’ and ‘quality of teaching’ in academic year 2012/2013;
- Binary variable for latest Ofsted rating for ‘leadership and management’ at least ‘good’ in academic year 2012/2013, as Achieve Together schools have particularly high scores on this scale;
- Continuous variable for percentage of pupils eligible for free school meals taken from the publicly available school-level census in academic year 2012/2013 (pnumfsm);
- Continuous variable for percentage of pupils with English as an additional language taken from the publicly available school-level census in academic year 2012/2013 (pnumeal);
- Continuous variable for percentage of pupils with SEN statement or on School Action Plus taken from the publicly available school-level census in academic year 2012/2013 (psensap).

Optimal propensity score matching specification: Nearest neighbour matching with six nearest neighbours and caliper of one. Independent matching variables used are:

- Continuous variable for % achieving 5A*-C GCSE and equivalents including English and maths in academic year 2012/2013 from the publicly available performance profiles (ptac5em);
- Continuous variable for total average (capped) point score per pupil (GCSEs only) in academic year 2012/2013 from the publicly available performance profiles (ttapsgcp);
- Continuous variable for value added measure based on the best 8 GCSE and equivalent results in academic year 2012/2013 from the publicly available performance profiles (b8vamea);
- Continuous variable for percentage of pupils with English as an additional language taken from the publicly available school-level census in academic year 2012/2013 (pnumeal);
- Binary variable for latest Ofsted rating for 'leadership and management' at least 'good' in academic year 2012/2013;

The exact code for the propensity score matching specification in Stata 14 is: `psmatch2 t2_2 fiveac_2013 gcseav_2013 va_best8_2013 peal_census_2013 leadership_good_2013, neighbor(6) cal(1)` where t2_2 is equal to one if the school is in the intervention group and zero if the school is in the potential set of comparison schools (who were eligible for Achieve Together and expressed some interest).

Analysis code is given below.

```
// create program to count schools in control and treatment group in each sample
cap program drop count_schools
program define count_schools
syntax, p_group(varlist)

        tempvar sample_temp
        qui gen `sample_temp' = e(sample)
        local N_schools = e(N_clust)

        tempvar n_school_0
        qui egen `n_school_0' = tag(ks4_urn) if `p_group' == 0 & `sample_temp' == 1
        qui count if `n_school_0' == 1
        local N_schools_0 = r(N)

        tempvar n_school_1
        egen `n_school_1' = tag(ks4_urn) if `p_group' == 1 & `sample_temp' == 1
        qui count if `n_school_1' == 1
        local N_schools_1 = r(N)
        assert `N_schools_0' + `N_schools_1' == `N_schools'

        qui estadd local N_schools `N_schools'
        qui estadd local N_schools_0 `N_schools_0'
        qui estadd local N_schools_1 `N_schools_1'

end


// Main effects

local base_year = 2013
local pupilx sob_2 sob_3 sob_4 ks4_male everfsm_2 everfsm_3 idaciscore_s idaciscore_s_miss
ks2_totpts_q_1 ks2_totpts_q_2 ks2_totpts_q_3 ks2_totpts_q_5
```

```
local schoolx fiveac_`base_year' gcseav_`base_year' va_best8_`base_year' qts_`base_year'
ofsted_average_`base_year' leadership_good_`base_year' pfsm_census_`base_year'
peal_census_`base_year' psen_census_`base_year'
local outcome ks4_ptscnewe_ptq_ee

eststo clear
discard
foreach npd_year in 2015 2016 {

        use "$pdata\data5_`npd_year'.dta", clear

        foreach outcome in `outcome' {

                qui tab `outcome'
                assert r(r) > 2 // assert continuous variable

                qui eststo est: reg `outcome' t2_2 if inlist(t2_2,0,1) & t2_2_sample == 1 &
t2_2_support == 1 [pweight=t2_2_weight], cluster(ks4_urn) robust
                count_schools, p_group(t2_2)

                esttab est using "$ptables/an7_`outcome'_`npd_year'.csv", label nogap se abs
bracket replace mtitle("Weighted + controls") stats(N N_schools N_schools_0 N_schools_1)

                }

        }
```

Interim analyses

There will be no interim analysis as the optimal matching specification has already been decided. This was completed in May 2016 using baseline characteristics of schools (from the 2012/2013 academic year). The optimal matching specification was chosen according to the specification with the lowest absolute percent median bias across key school variables (from the 2012/2013 academic year). The matching specifications tested varied the independent covariates and propensity score matching method. All specifications tested are shown in Appendix 1.

Imbalance at baseline

There will be no school-level attrition from the evaluation as the NPD will be used to access pupil-level attainment and absence data for all schools that signed up to the trial and all potential comparison schools. Attrition will therefore not lead to imbalance at baseline (between the intervention and potential comparison group) and is therefore not included in the evaluation design.

Missing data

Missing data will not be imputed. Pupils with missing values for at least one primary and secondary outcome (aside from absence) will be excluded. Based on previous years' data under 3% of KS4 pupils would be excluded using this criterion. The absence variable has more missing values (around 10% of the sample in previous years), and so the common sample will not be restricted to non-missing values of this variable. Sensitivity analysis that restricts the common sample to include non-missing values of all primary and secondary outcomes will be presented in an appendix.

Missing values of independent variables will be accounted for by binary variables. These binary variables will capture the differences in outcome variable for those with and without missing values, on average, which are likely to be driven by the non-random selection. For example, missing KS2 values are likely to reflect not being present in the state sector in England at the end of primary school. (Missing KS2 values are more likely for the KS4 cohort

in 2015 – after the first year of Achieve Together – due to the KS2 boycott in 2010. Missing values for this cohort due to school-level boycott will be imputed using the following variables from the NPD Tables: KS4_VAP2TAENG_PTQ and KS4_VAP2TAMAT_PTQ. An binary variable equal to one if these imputed variables are used and zero if not will be included. Sensitivity analysis will be conducted using these variables for all students.)

On-treatment analysis

On treatment will be defined as having at least one participant from the school on the senior leadership training programme (NPQSL) over the two year programme. The analysis model will be identical but with the four schools that do not meet this criterion excluded.

Secondary outcome analyses

The secondary outcome analysis will exactly replicate the primary outcome analysis, but using each secondary outcome variable as the dependent variable (in turn) rather than the primary outcome variable.

The only addition is that for the absence secondary outcome variable the individual-level continuous measure of absence from the previous academic year will be used as an independent variable.

Additional analyses

- MDES calculation – on the basis of actual parameters seen
- School characteristics of intervention schools
- School characteristics of comparison schools, before and after propensity score matching is used to define the most well-matched comparison group of schools.
- Equivalent exercise for Achieve Together schools in the previous cohort, to inform whether the well-matched comparison group method is likely to be unbiased. This will be done by using the second cohort as a comparison group for the first cohort, and comparing this impact estimate to the estimate from the well-matched comparison group. This analysis will be presented in an appendix.

Subgroup analyses

Sub-group analysis on the primary outcome will be carried out on the following groups only as per the protocol: whether or not a pupil has ever received free school meals (everFSM); pupils with low prior attainment. Low prior attainment is defined as not achieving at least level 4 at Key Stage 2 in each of maths, English and (teacher assessed) science. This will be done using a model identical to the primary outcome model but splitting the sample to include the subgroup of interest only.

Analysis code is given below.

```
// FSM subgroup

local base_year = 2013
local pupilx sob_2 sob_3 sob_4 ks4_male idaciscore_s idaciscore_s_miss ks2_totpts_q_1
ks2_totpts_q_2 ks2_totpts_q_3 ks2_totpts_q_5
local schoolx fiveac_`base_year' gcseav_`base_year' va_best8_`base_year' qts_`base_year'
ofsted_average_`base_year' leadership_good_`base_year' pfsm_census_`base_year'
peal_census_`base_year' psen_census_`base_year'
```

```
local outcome ks4_ptscnewe_ptq_ee

eststo clear
discard
foreach npd_year in 2015 2016 {

        use "$pdata\data5_`npd_year'.dta", clear
        keep if everfsm_2 == 1

        foreach outcome in `outcome' {

                qui tab `outcome'
                assert r(r) > 2 // assert continuous variable

                qui eststo est: reg `outcome' t2_2 if inlist(t2_2,0,1) & t2_2_sample == 1 &
t2_2_support == 1 [pweight=t2_2_weight], cluster(ks4_urn) robust
                count_schools, p_group(t2_2)

                esttab est using "$ptables/an7_`outcome'_`npd_year'_fsm.csv", label nogap se
abs bracket replace mtitle("Weighted + controls") stats(N N_schools N_schools_0 N_schools_1)

                }

        }


// Low attainment subgroup

local base_year = 2013
local pupilx sob_2 sob_3 sob_4 ks4_male everfsm_2 everfsm_3 idaciscore_s idaciscore_s_miss
ks2_totpts_q_1 ks2_totpts_q_2 ks2_totpts_q_3 ks2_totpts_q_5
local schoolx fiveac_`base_year' gcseav_`base_year' va_best8_`base_year' qts_`base_year'
ofsted_average_`base_year' leadership_good_`base_year' pfsm_census_`base_year'
peal_census_`base_year' psen_census_`base_year'
local outcome ks4_ptscnewe_ptq_ee ks4_gptscnewe_ptq_ee ks4_engmatsci ks4_teachfirst
overallabsence_6halfterms_ab

eststo clear
discard
foreach npd_year in 2015 2016 {

        use "$pdata\data5_`npd_year'.dta", clear
        keep if ks2_lowattainment == 1

        foreach outcome in `outcome' {

                qui tab `outcome'
                assert r(r) > 2 // assert continuous variable

                qui eststo est: reg `outcome' t2_2 if inlist(t2_2,0,1) & t2_2_sample == 1 &
t2_2_support == 1 [pweight=t2_2_weight], cluster(ks4_urn) robust
                count_schools, p_group(t2_2)

                esttab est using "$ptables/an7_`outcome'_`npd_year'_low.csv", label nogap se
abs bracket replace mtitle("Weighted + controls") stats(N N_schools N_schools_0 N_schools_1)

                }

        }
```

Effect size calculation

Effect sizes will be calculated using multivariate regression, where the dependent variable is standardised according to the national distribution to have a mean of zero and standard deviation of one. Weights will be used to reflect the propensity score matching specification that creates the best balance between intervention and matched comparison group. Effect sizes will be reported with a 95% confidence interval that takes into account the clustered nature of the data.

Report tables

Table 1: Summary of impact on primary outcome

Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	EEF cost rating
Treatment 1 vs. Matched comparison group				
Treatment 2 vs. Matched comparison group				
Treatment 1 FSM vs. Matched comparison group				
Treatment 2 FSM vs. Matched comparison group				

Table 2: Minimum detectable effect size at different stages

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol							
Recruited							
Analysis							

Table 3: Baseline comparison (unmatched)

Variable	Intervention group		Potential comparison group	
School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Ofsted overall effectiveness outstanding				

or good				
Ofsted quality of teaching outstanding or good				
Ofsted leadership and management outstanding or good				
School-level (continuous)	n (missing)	[Mean or median]	n (missing)	[Mean or median]
% 5A*-C (including English and maths)				
Capped points score				
Value added				
Teachers with QTS (%)				
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for FSM				
English as an additional language				
Special Educational Need				

Table 4: Baseline comparison (matched)

Variable	**Intervention group**		**Potential comparison group**	
School-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Ofsted overall effectiveness outstanding or good				
Ofsted quality of teaching outstanding or good				
Ofsted leadership and management outstanding or good				
School-level (continuous)	n (missing)	[Mean or median]	n (missing)	[Mean or median]
% 5A*-C (including English and maths)				
Capped points score				
Value added				
Teachers with QTS (%)				

Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for FSM				
English as an additional language				
Special Educational Need				

Table 5: Primary analysis

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS4 points score (2016)							

Table 6: Secondary analysis (end of year two of Achieve Together)

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS4 points score (excluding equivalents) (2016)							
Teach First impact goal measure (2016)							
Overall absence (2016)							

Table 7: Secondary analysis (end of year one of Achieve Together)

Raw means	Effect size

Outcome	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS4 points score (2015)							
KS4 points score (excluding equivalents) (2015)							
Teach First impact goal measure (2015)							
Overall absence (2015)							

Table 8: Primary analysis (on-treatment)

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS4 points score (2016)							

Appendix Table 1: Sensitivity analysis (end of year two of Achieve Together imposing a common sample for all outcomes)

Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS4 points score (2016)							

KS4 points score (excluding equivalents) (2016)							
Teach First impact goal measure (2016)							
Overall absence (2016)							

Appendix Table 2: Sensitivity analysis (including school-level random effects)

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**n in model (intervention; control)**	**Hedges g (95% CI)**	**p-value**
KS4 points score (2016)							
KS4 points score (excluding equivalents) (2016)							
Teach First impact goal measure (2016)							
Overall absence (2016)							

Appendix Table 3: Secondary analysis (year one estimates for Achieve Together cohort 1)

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n**	**Mean**	**n in model (intervention;**	**Hedges g (95% CI)**	**p-value**

			(missing)	(95% CI)	control)		
KS4 points score (2015)							
KS4 points score (excluding equivalents) (2015)							
Teach First impact goal measure (2015)							
Overall absence (2015)							

Appendix 1

Appendix 1 contains the description and syntax for finding the optimal matching specification. In summary, the optimal matching specification was chosen by finding the specification with the lowest absolute median bias across the following variables:

- Continuous variable for % achieving 5A*-C GCSE and equivalents including English and maths in academic year 2012/2013 from the publicly available performance profiles (ptac5em);
- Continuous variable for total average (capped) point score per pupil (GCSEs only) in academic year 2012/2013 from the publicly available performance profiles (ttapsgcp);
- Continuous variable for value added measure based on the best 8 GCSE and equivalent results in academic year 2012/2013 from the publicly available performance profiles (b8vamea). Value added, total average (capped) points score and %5A*-C have a relatively low correlation for the group of intervention schools which justifies inclusion of all three: between 0.34 and 0.88;
- Continuous variable for % teachers with Qualified Teacher Status in academic year 2012/2013 from publicly available school-level data derived from the School Workforce Census (qts). The theoretical reasoning for this variable is that schools are likely to have signed up to Achieve Together to address staff shortages, which are proxied by this variable;
- Continuous variable for average of latest Ofsted rating for 'overall effectiveness', 'leadership and management' and 'quality of teaching' in academic year 2012/2013;
- Binary variable for latest Ofsted rating for 'leadership and management' at least 'good' in academic year 2012/2013, as Achieve Together schools have particularly high scores on this scale;
- Continuous variable for percentage of pupils eligible for free school meals taken from the publicly available school-level census in academic year 2012/2013 (pnumfsm);
- Continuous variable for percentage of pupils with English as an additional language taken from the publicly available school-level census in academic year 2012/2013 (pnumeal);
- Continuous variable for percentage of pupils with SEN statement or on School Action Plus taken from the publicly available school-level census in academic year 2012/2013 (psensap).

These variables were chosen through theory as the variables that are most likely to represent differences in schools' motivation to sign up to Achieve Together (or not) and pupil outcomes.

Various propensity score matching methods were tested, varying the variables and tolerance used for each method. These methods are:

- Kernel: varying bandwidth and variables used to create the propensity score;

- Neighest neighbour(s): varying the caliper and variables used to create the propensity score;
- Mahalanobis: varying the variables used to create the Mahalanobis distance.

```
*      PROGRAM to test matching specifications

cap program drop at_matching

program define at_matching


version 11

syntax [if] [in], treatment(string) matchingvariables(varlist) writeto(string)
acyear(integer) [psmatchoptions(string) pscore]


       tempname file

       file open `file' using `"`writeto'"', write text replace

       file write `file' _n

       if "`treatment'" == "t1" file write `file' "Alternative control groups for
cohort 1" _n

       if "`treatment'" == "t2" file write `file' "Alternative control groups for
cohort 2" _n

       if "`psmatchoptions'" == "" file write `file' "Unweighted" _n

       else if "`psmatchoptions'" != "" file write `file' "Weighted:
`psmatchoptions'" _n

       if "`treatment'" == "t1" file write `file' "Variable," "Cohort 1," "All
eligible schools (Cohort 1)," "Eigible schools that expressed interest (Cohort 1),"
"Cohort 2"   _n

       if "`treatment'" == "t2" file write `file' "Variable," "Cohort 2," "All
eligible schools (Cohort 2)," "Eligible schools that expressed interest (Cohort 2)"
_n


       local vargroup1 fiveac_`acyear' gcseav_`acyear' va_best8_`acyear'

       local vargroup2 qts_`acyear'

       local vargroup3 /*effectiveness_`acyear' teaching_`acyear'
leadership_`acyear'*/ ofsted_average_`acyear' leadership_good_`acyear'

       local vargroup4 pfsm_census_`acyear' peal_census_`acyear'
psen_census_`acyear'

       if "`pscore'" != "" {

              if "`treatment'" == "t1" local pscorevar pscore_t1

              if "`treatment'" == "t2" local pscorevar pscore_t2

              }

       else if "`pscore'" == "" local pscorevar
```

```
	discard
	cap drop weight_*

	assert "`treatment'" == "t1" | "`treatment'" == "t2"
	if "`treatment'" == "t1" local groupstart = 1
	else if "`treatment'" == "t2" local groupstart

	foreach group in 4 2 `groupstart' {

		psmatch2 `treatment'_`group' `matchingvariables' , `psmatchoptions'
		rename _weight weight_`group'
		qui sum `vargroup1' `vargroup2' `vargroup3' `vargroup4' if
inlist(`treatment'_`group',0,1)
		pstest_3 `vargroup1' `vargroup2' `vargroup3' `vargroup4',
treated(`treatment'_`group') both mweight(weight_`group')
		local unw_`group' = `r(medbiasbef)'
		local w_`group' = `r(medbiasaft)'
		local B_unw_`group' = `r(Bbef)'
		local B_w_`group' = `r(Baft)'
		if `group' == 4 {
			qui sum interested [aweight=weight_4] if `treatment'_4 == 0
			local prop_interested = r(mean)
			}
		forvalues vargroup = 1/4 {
			qui pstest_3 `vargroup`vargroup'', treated(`treatment'_`group')
both mweight(weight_`group')
			local unw_`group'_`vargroup' = `r(medbiasbef)'
			local w_`group'_`vargroup' = `r(medbiasaft)'
			}

		tab `treatment'_`group' _support, miss
		}



	foreach var of varlist fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' ///
		qts_`acyear' ///
		ofsted_average_`acyear' leadership_good_`acyear' ///
```

```
		pfsm_census_`acyear' peal_census_`acyear' psen_census_`acyear' {


		qui sum `var' if `treatment'_2 == 1

		local mean_c1 = r(mean)


		foreach group in 4 2 `groupstart' {


			qui summarize `var' if `treatment'_`group' == 0
[aweight=weight_`group']   // loop over alternative control groups

			local mean`group' = r(mean)


			}


		if "`var'" == "fiveac_`acyear'" file write `file' "Attainment" _n

		if "`var'" == "qts_`acyear'" file write `file' "School workforce" _n

		if "`var'" == "ofsted_average_`acyear'" file write `file' "Ofsted" _n

		if "`var'" == "pfsm_census_`acyear'" file write `file' "Pupil
composition" _n


		if "`treatment'" == "t1" {

			file write `file' "`: variable label `var'," %6.2f (`mean_c1')
"," %6.2f (`mean4') "," %6.2f (`mean2') "," %6.2f (`mean1') _n

			if "`var'" == "va_best8_`acyear'" {     // last in group so
report standardised bias (academic attainment)

				file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_1') "," %6.2f (`unw_2_1') "," %6.2f (`unw_1_1') _n

				file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_1') "," %6.2f (`w_2_1') "," %6.2f (`w_1_1') _n

				}

			if "`var'" == "qts_`acyear'" {    // (school staff comp)

				file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_2') "," %6.2f (`unw_2_2') "," %6.2f (`unw_1_2') _n

				file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_2') "," %6.2f (`w_2_2') "," %6.2f (`w_1_2') _n

				}

			if "`var'" == "leadership_good_`acyear'" {     // (Ofsted)

				file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_3') "," %6.2f (`unw_2_3') "," %6.2f (`unw_1_3') _n

				file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_3') "," %6.2f (`w_2_3') "," %6.2f (`w_1_3') _n

				}
```

```
                    if "`var'" == "psen_census_`acyear'" {  // (school pupil comp)

                            file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_4') "," %6.2f (`unw_2_4') "," %6.2f (`unw_1_4') _n

                            file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_4') "," %6.2f (`w_2_4') "," %6.2f (`w_1_4') _n

                            file write `file' "Total: Raw |% standardised bias|, ,"
%6.2f (`unw_4') "," %6.2f (`unw_2') "," %6.2f (`unw_1') _n

                            file write `file' "Total: Weighted |% standardised
bias|, ," %6.2f (`w_4') "," %6.2f (`w_2') "," %6.2f (`w_1') _n

                            file write `file' "Total: B before, ," %6.2f (`B_unw_4')
"," %6.2f (`B_unw_2') "," %6.2f (`B_unw_1') _n

                            file write `file' "Total: B after, ," %6.2f (`B_w_4')
"," %6.2f (`B_w_2') "," %6.2f (`B_w_1') _n

                            file write `file' "Weighted prop. of elig. schools
interested, ," %6.2f (`prop_interested')

                            }

                    }


            if "`treatment'" == "t2" {


                    file write `file' "`: variable label `var'," %6.2f (`mean_c1')
"," %6.2f (`mean4') "," %6.2f (`mean2') _n

                    if "`var'" == "va_best8_`acyear'" {     // last in group so
report standardised bias (academic attainment)

                            file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_1') "," %6.2f (`unw_2_1') _n

                            file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_1') "," %6.2f (`w_2_1') _n

                            }

                    if "`var'" == "qts_`acyear'" {    // (school staff comp)

                            file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_2') "," %6.2f (`unw_2_2') _n

                            file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_2') "," %6.2f (`w_2_2') _n

                            }

                    if "`var'" == "leadership_good_`acyear'" {     // (Ofsted)

                            file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_3') "," %6.2f (`unw_2_3') _n

                            file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_3') "," %6.2f (`w_2_3') _n

                            }

                    if "`var'" == "psen_census_`acyear'" {  // (school pupil comp)

                            file write `file' "Raw |% standardised bias|, ," %6.2f
(`unw_4_4') "," %6.2f (`unw_2_4') _n
```

```
                        file write `file' "Weighted |% standardised bias|, ,"
%6.2f (`w_4_4') "," %6.2f (`w_2_4') _n

                        file write `file' _n

                        file write `file' "Total: Raw |% standardised bias|, ,"
%6.2f (`unw_4') "," %6.2f (`unw_2') _n

                        file write `file' "Total: Weighted |% standardised
bias|, ," %6.2f (`w_4') "," %6.2f (`w_2') _n

                        file write `file' "Total: B before, ," %6.2f (`B_unw_4')
"," %6.2f (`B_unw_2') _n

                        file write `file' "Total: B after, ," %6.2f (`B_w_4')
"," %6.2f (`B_w_2') _n

                        file write `file' "Weighted prop. of elig. schools
interested, ," %6.2f (`prop_interested') _n

                        }

                    }

            }


      file write `file' "Note: Matching spec: `psmatchoptions'. Matching
variables: `matchingvariables'" _n


      end


**********************




*     use program for Cohort 2

use "$pdata\other_data_1404.dta", clear


discard

at_matching, treatment(t2) matchingvariables(avpoints_2013 fiveac_2013
ftvacant_2013 temps_2013 qts_2013 tatratio_2013 sen_2013 eal_2013 fsm_2013)
acyear(2013) psmatchoptions(neighbor(5)) ///

      writeto("$ptables\an4_t2_test_v1.csv")

// start with least balanced characteristics only -> reasonable, but poor balance
on pupil comp for eligible group 7.69, 14.87

at_matching, treatment(t2) matchingvariables(va_best8_2013 fiveac_2013 eal_2013
ofsted_average_2013 leadership_good_2013) acyear(2013) psmatchoptions(neighbor(5))
///

      writeto("$ptables\an4_t2_test_v2.csv")

// account for five AC attainment -> better overall and noticeably better for
attainment for eligble group 6.93, 7.87
```

```
at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(5)) ///

       writeto("$ptables\an4_t2_test_v3.csv")

// add extra Ofsted indicators -> doesn't improve Ofsted match and makes other
things worse -> 13.25, 14.83

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013 effectiveness_2013)
acyear(2013) psmatchoptions(neighbor(5)) ///

       writeto("$ptables\an4_t2_test_v4.csv")

// add QTS -> better match on QTS but worse overall -> 17.55, 14.12

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(neighbor(5)) ///

       writeto("$ptables\an4_t2_test_v5.csv")

// add FSM -> better than above but not best -> 14.69, 8.85

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 fsm_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(neighbor(5)) ///

       writeto("$ptables\an4_t2_test_v6.csv")

// add caliper -> no difference to best spec (v3) -> 6.93, 7.87

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(5) cal(1)) ///

       writeto("$ptables\an4_t2_test_v7.csv")

// decrease NN -> worse -> 17.49, 13.36

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(2) cal(1)) ///

       writeto("$ptables\an4_t2_test_v8.csv")

// increase NN -> much better for eligible schools -> 1.81, 14.78

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(10) cal(1)) ///

       writeto("$ptables\an4_t2_test_v9.csv")

// increase NN and add qts -> better for QTS but worse overall -> 7.31, 9.14

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(neighbor(10) cal(1)) ///

       writeto("$ptables\an4_t2_test_v10.csv")

// increase NN and add fsm -> better for QTS but worse overall -> 13.15, 9.27

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 fsm_2013 eal_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(neighbor(10) cal(1)) ///

       writeto("$ptables\an4_t2_test_v11.csv")
```

```
// experiment with NN number - 10 is the best

forvalues n = 1/20 {

	at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(`n') cal(1) qui) ///

		writeto("$ptables\an4_t2_test_v9_`n'.csv")

	}



// best NN spec with Mahalonbis -> worse -> 20, 10.65

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(mahalanobis(avpoints_2013 va_best8_2013 fiveac_2013 eal_2013
ofsted_average_2013 leadership_good_2013)) ///

	writeto("$ptables\an4_t2_test_v12.csv")

// add pscore -> worse -> 27.58, 30.35

reg t2_4 avpoints_2013 va_best8_2013 fiveac_2013 qts_2013 eal_2013 fsm_2013
ofsted_average_2013 leadership_good_2013

predict pscore_4

reg t2_2 avpoints_2013 va_best8_2013 fiveac_2013 qts_2013 eal_2013 fsm_2013
ofsted_average_2013 leadership_good_2013

predict pscore_2

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(mahalanobis(pscore_2)) ///

	writeto("$ptables\an4_t2_test_v13.csv")

drop pscore_*

// more limited set of variables for Mahal -> better for interested group -> 8.2,
4.07

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(mahalanobis(va_best8_2013 fiveac_2013 eal_2013
leadership_2013)) ///

	writeto("$ptables\an4_t2_test_v14.csv")



// Kernel - start with best NN spec - worse -> 11.92, 17.45

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(kernel) ///

	writeto("$ptables\an4_t2_test_v15.csv")

// Kernel - add QTS -> better -> 10.19, 11.11

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(kernel) ///
```

```
	writeto("$ptables\an4_t2_test_v16.csv")

// Kernel - add FSM -> better -> 9.25, 8.27

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 fsm_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(kernel) ///

	writeto("$ptables\an4_t2_test_v17.csv")

// Kernel - try smaller bandwidth - similar -> 6.63, 10.71

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 fsm_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(kernel bwidth(0.03)) ///

	writeto("$ptables\an4_t2_test_v18.csv")

// Kernel - try bigger bandwidth - similar -> 7.91, 7.53

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 fsm_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(kernel bwidth(0.10)) ///

	writeto("$ptables\an4_t2_test_v19.csv")

// experiment with bandwidth - 0.09 is best for interested group (7.27)

local test = 1

forvalues band = 0.01(0.01)0.20 {

	local test = `test' + 1

	at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 fsm_2013 ofsted_average_2013 leadership_good_2013)
acyear(2013) psmatchoptions(kernel bwidth(`band') qui) ///

		writeto("$ptables\an4_t2_test_v19_`test'.csv")

	}



// Kernel - drop ofsted extra variables -> similar -> 7.62, 8.43

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 fsm_2013 leadership_good_2013) acyear(2013)
psmatchoptions(kernel bwidth(0.09)) ///

	writeto("$ptables\an4_t2_test_v20.csv")

// Kernel - leadership and effectiveness -> similar -> 7.81, 7.75

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 fsm_2013 leadership_good_2013 effectiveness_2013)
acyear(2013) psmatchoptions(kernel bwidth(0.09)) ///

	writeto("$ptables\an4_t2_test_v21.csv")

// Kernel - effectiveness -> similar -> 9.13, 7.28

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013
fiveac_2013 qts_2013 eal_2013 effectiveness_2013) acyear(2013)
psmatchoptions(kernel bwidth(0.09)) ///

	writeto("$ptables\an4_t2_test_v22.csv")

// Kernel - drop five ac -> similar -> 10.02, 7.21
```

```
at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013 qts_2013
eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(kernel bwidth(0.09)) ///

       writeto("$ptables\an4_t2_test_v23.csv")

// Kernel - add FSM -> remains reasonable for eligible group -> 10.02, 8.38

at_matching, treatment(t2) matchingvariables(avpoints_2013 va_best8_2013 qts_2013
fsm_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(kernel bwidth(0.09)) ///

       writeto("$ptables\an4_t2_test_v24.csv")



// work on NN spec for the interested group -> better match of pupil comp but worse
overall and elsewhere -> 15.19, 13.24

at_matching, treatment(t2) matchingvariables(va_best8_2013 fiveac_2013 eal_2013
fsm_2013 sen_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(5)) ///

       writeto("$ptables\an4_t2_test_v2_b.csv")

// work on NN spec for the interested group -> better match of pupil comp but worse
overall and elsewhere -> 15.49, 13.07

at_matching, treatment(t2) matchingvariables(va_best8_2013 fiveac_2013 eal_2013
fsm_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(5)) ///

       writeto("$ptables\an4_t2_test_v2_c.csv")

// work on NN spec for the interested group -> 7.55, 13.12

at_matching, treatment(t2) matchingvariables(va_best8_2013 fiveac_2013
avpoints_2013 eal_2013 ofsted_average_2013 leadership_good_2013) acyear(2013)
psmatchoptions(neighbor(5)) ///

       writeto("$ptables\an4_t2_test_v2_d.csv")

// work on NN spec for the interested group -> 7

at_matching, treatment(t2) matchingvariables(va_best8_2013 fiveac_2013 eal_2013
ofsted_average_2013 leadership_good_2013) acyear(2013) psmatchoptions(neighbor(6))
///

       writeto("$ptables\an4_t2_test_v2_e.csv")



*      check Mahal code - best is still median bias of 14.9

cap drop pscore

logit t2_2 avpoints_2013 va_best8_2013 fiveac_2013 qts_2013 eal_2013 fsm_2013
sen_2013 ofsted_average_2013 leadership_good_2013

predict pscore

psmatch2 t2_2, mahalanobis(pscore) bwidth(0.01)

local acyear 2013
```

```
pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' fsm_`acyear' eal_`acyear'
sen_`acyear'


cap drop pscore

logit t2_2 avpoints_2013 va_best8_2013 fiveac_2013 qts_2013 eal_2013 fsm_2013
sen_2013 ofsted_average_2013 leadership_good_2013

predict pscore

psmatch2 t2_2, mahalanobis(pscore) bwidth(0.10)

local acyear 2013

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' fsm_`acyear' eal_`acyear'
sen_`acyear'


cap drop pscore

logit t2_2 va_best8_2013 eal_2013 leadership_good_2013 gcseav_2013

predict pscore

psmatch2 t2_2, mahalanobis(pscore) bwidth(0.10)

local acyear 2013

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' fsm_`acyear' eal_`acyear'
sen_`acyear'


**********************************************************************

*Now don't use program to explore matching - iterate based on pstest

**********************************************************************

* NOTE: use composition from census rather than attainment tables


*      Cohort 2 - interested

***************************

local acyear 2013

psmatch2 t2_2 peal_census_`acyear', kernel

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss


// most improved. Add VA, five AC, capped points score

local acyear 2013

psmatch2 t2_2 peal_census_`acyear' fiveac_`acyear' gcseav_`acyear'
va_best8_`acyear', kernel
```

```
pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss


// Add QTS - OK -> 8.7

local acyear 2013

psmatch2 t2_2 peal_census_`acyear' fiveac_`acyear' gcseav_`acyear'
va_best8_`acyear' qts_`acyear', kernel

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss


// Add leadership

local acyear 2013

psmatch2 t2_2 peal_census_`acyear' fiveac_`acyear' gcseav_`acyear'
va_best8_`acyear' qts_`acyear' leadership_good_`acyear', kernel

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss


// NNs -> bias 8.2

local acyear 2013

psmatch2 t2_2 fiveac_`acyear' gcseav_`acyear' va_best8_`acyear'
peal_census_`acyear', neighbor(5) cal(1)

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss


// Add QTS -> bias 11.2

local acyear 2013

psmatch2 t2_2 fiveac_`acyear' gcseav_`acyear' va_best8_`acyear'
peal_census_`acyear' qts_`acyear', neighbor(5) cal(1)

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss


// Add leadership -> bias 3.9

local acyear 2013
```

```
psmatch2 t2_2 fiveac_`acyear' gcseav_`acyear' va_best8_`acyear'
peal_census_`acyear' leadership_good_`acyear', neighbor(5) cal(1)

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss



// Add SEN -> bias 10.3

local acyear 2013

psmatch2 t2_2 fiveac_`acyear' gcseav_`acyear' va_best8_`acyear'
peal_census_`acyear' leadership_good_`acyear' psen_census_`acyear', neighbor(5)
cal(1)

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss



// experiment with neighbours -> bias 2.5 when N == 6

local acyear 2013

psmatch2 t2_2 fiveac_`acyear' gcseav_`acyear' va_best8_`acyear'
peal_census_`acyear' leadership_good_`acyear', neighbor(6) cal(1)

pstest fiveac_`acyear' gcseav_`acyear' va_best8_`acyear' qts_`acyear'
ofsted_average_`acyear' leadership_good_`acyear' pfsm_census_`acyear'
peal_census_`acyear' psen_census_`acyear', both

tab _support t2_2, miss
```